

Analisis *Clustering* Metode Hierarki *Single Linkage* terhadap Indeks Pembangunan Manusia (IPM) Indonesia Tahun 2023

Clustering Analysis of Single Linkage Hierarchical Method on Indonesia's Human Development Index (IPM) in 2023

Ardita Rahastri Pertiwi¹, Farah Madina², Fika Yulistina³, Aisy Rohmah Najahy Mawaddah⁴, Affiati Oktaviarina⁵

^{1,2,3,4,5}State University of Surabaya, Faculty off Mathematics and Natural Sciences

Article Info:

Received: 17-12-2023

in revised form: 23-12-2023

Accepted: 04-04-2024

Available Online: 10-04-2024

Keywords:

IPM, clustering, single linkage

Corresponding Author:

affiatioktaviarina@unesa.ac.id

ardita.20032@mhs.unesa.ac.id

id;

Abstract: *The success of a country's or region's development heavily relies on the quality of its population, which can be measured through the Human Development Index (HDI). Although Indonesia's HDI increased by 0.62 points in 2023, this increment does not ensure uniform prosperity across society. Therefore, it is necessary to classify provinces in Indonesia based on HDI data to assist the government in designing more precisely targeted development programs. This research utilizes the single linkage hierarchical cluster analysis method, and the results indicate a division into 4 clusters with distinct characteristics. Cluster 1 comprises 31 provinces with favorable characteristics, Cluster 2 consists of DI Yogyakarta with favorable characteristics, Cluster 3 consists of DKI Jakarta with less favorable characteristics, and Cluster 4 consists of Papua with very unfavorable characteristics. The findings of this study demonstrate that the single linkage hierarchical method is effective in grouping provinces based on HDI, with a high cophenetic correlation coefficient of 0.83.*

Abstrak: Keberhasilan pembangunan suatu negara atau wilayah sangat bergantung pada kualitas penduduknya, yang dapat diukur melalui Indeks Pembangunan Manusia (IPM). Meskipun IPM Indonesia meningkat sebesar 0,62 poin pada tahun 2023, kenaikan ini tidak menjamin kesejahteraan yang merata di seluruh masyarakat. Oleh karena itu, diperlukan pengelompokan provinsi di

Indonesia berdasarkan data IPM untuk membantu pemerintah merancang program pembangunan yang lebih tepat sasaran. Penelitian ini menggunakan metode analisis *cluster* hierarki *single linkage* dan hasilnya menunjukkan pembagian menjadi 4 cluster yang berbeda karakteristik. Klaster 1 terdiri dari 31 provinsi dengan karakteristik baik, Klaster 2 terdiri dari DI Yogyakarta dengan

karakteristik baik, Klaster 3 terdiri dari DKI Jakarta dengan karakteristik kurang, dan Klaster 4 terdiri dari Papua dengan karakteristik sangat kurang. Hasil penelitian ini menunjukkan bahwa metode hierarki *single linkage* efektif dalam pengelompokan provinsi berdasarkan IPM, dengan nilai koefisien korelasi *cophenetic* yang tinggi, yaitu 0,83.

I. PENDAHULUAN

Keberhasilan dari pembangunan sebuah negara atau wilayah tidak terlepas dari peran penting penduduknya. Jika penduduk di suatu daerah terjamin kualitasnya, itu akan menjadi modal dalam pembangunannya. Namun jika kualitas dari penduduk tersebut rendah, justru hal ini akan menjadikan sebuah beban dalam pembangunan.

Pada perencanaan pembangunan suatu negara atau wilayah, indikator yang dibutuhkan salah satunya adalah Indeks Pembangunan Manusia (IPM). Badan Pusat Statistik (BPS) Indonesia menyatakan bahwa Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup serta menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. Tingkat kesejahteraan dan kualitas hidup penduduk suatu negara atau wilayah dapat dicerminkan oleh data IPM itu sendiri, sehingga IPM ini juga berperan sebagai penentu peringkat atau level pembangunan suatu negara atau wilayah (Wulansari, 2012).

Menurut Badan Pusat Statistik (BPS) Indonesia, IPM dibangun melalui pendekatan tiga dimensi dasar yaitu umur panjang dan kesehatan, pengetahuan, serta standar kehidupan yang layak. Ketiga dimensi tersebut dibagi ke dalam 4 komponen yaitu Umur Harapan Hidup Saat Lahir (AHH), Harapan Lama Sekolah (HLS), Pengeluaran per Kapita Disesuaikan (PKD), dan Rata - Rata Lama Sekolah (RLS) (Widodo, Mashita, & Prasetyowati, 2020).

Pada tahun 2023, Indeks Pembangunan Manusia (IPM) Indonesia meningkat 0,62 poin dari tahun sebelumnya sebesar 73,77 (Badan Pusat Statistik, 2023). Pertumbuhan IPM ini berarti menunjukkan adanya upaya dari Pemerintah Indonesia dalam pemulihan dan peningkatan kesejahteraan penduduk di Indonesia. Akan tetapi, mayoritas dari data IPM tahun 2023 yang diambil dari 34 Provinsi yang terdapat di Indonesia, masih banyak provinsi-provinsi yang memiliki perbedaan sangat jauh. Hal ini kemudian yang mengakibatkan pembangunan di setiap provinsi di Indonesia belum adil dan merata.

Pada tahun 2023, Indeks Pembangunan Manusia (IPM) Indonesia meningkat sebesar 0,62 poin dari tahun sebelumnya, mencapai 73,77 (Badan Pusat Statistik, 2023). Pertumbuhan ini mencerminkan upaya yang dilakukan oleh Pemerintah Indonesia dalam pemulihan dan peningkatan kesejahteraan penduduk. Namun, meskipun terjadi peningkatan secara keseluruhan, data IPM tahun 2023 menunjukkan ketimpangan yang signifikan antar provinsi di Indonesia. Ketimpangan ini mengindikasikan bahwa pembangunan di setiap provinsi belum merata dan adil.

Sebagai contoh, penelitian sebelumnya yang dilakukan oleh Matdoan, M. Y., & Noya Van Delsen, M. S. (2020), berjudul "Penerapan Analisis Cluster dengan Metode Hierarki untuk Klasifikasi Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia," bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan IPM dari tahun 2010 sampai 2019 menggunakan berbagai metode analisis klaster. Penelitian tersebut berhasil mengidentifikasi empat cluster dengan tingkat IPM yang berbeda. Namun, dalam penelitian ini, langkahnya lebih jauh dengan mengklasterkan tingkat IPM secara nasional. Perspektif yang lebih luas diangkat dengan menganalisis pola pengelompokan IPM di seluruh

Indonesia, bukan hanya di satu provinsi. Hal ini memberikan wawasan yang lebih komprehensif tentang ketimpangan pembangunan manusia di negara ini.

Dari permasalahan di atas, kemudian penulis tergugah untuk mempertimbangkan pengelompokan karakteristik provinsi di Indonesia berdasarkan data IPM. Proses pengelompokan provinsi menggunakan data IPM melibatkan analisis terhadap nilai IPM dari setiap provinsi, yang biasanya diperoleh dari Badan Pusat Statistik (BPS) atau lembaga terkait lainnya. Provinsi-provinsi dengan nilai IPM yang tinggi dapat dianggap memiliki tingkat pembangunan manusia yang lebih baik, sementara provinsi dengan nilai IPM yang rendah mungkin memerlukan perhatian lebih dalam upaya pembangunan. Pentingnya pengelompokan provinsi berdasarkan IPM adalah untuk membantu pemerintah beserta stafnya dalam merancang kebijakan pembangunan yang lebih tepat sasaran. Dengan mengetahui perbedaan tingkat pembangunan antar provinsi, pemerintah dapat merancang program pembangunan yang lebih spesifik dan efektif sesuai dengan kebutuhan setiap wilayah. Dalam menganalisis pengelompokan provinsi ini menggunakan metode analisis *cluster* hierarki yaitu *single linkage*. Menurut Alvin C. Rencher (2002), metode *single linkage* adalah salah satu metode pengelompokan data yang menggabungkan dua data yang memiliki jarak terkecil di antara mereka, sehingga membentuk satu klaster baru. Penggunaan metode *single linkage*, dipilih karena alasan kepraktisan dan kemudahan implementasi. Metode ini telah terbukti efektif dalam pengelompokan data serupa dalam riset sebelumnya, seperti yang dilakukan oleh Hanada & Yanti (2021) dalam pengelompokan kecamatan di Kabupaten Karawang berdasarkan metode kontrasepsi peserta KB aktif, Harnanto et al. (2017) dalam pengelompokan kabupaten/kota di Jawa Tengah berdasarkan pengguna alat kontrasepsi, serta penelitian-penelitian lainnya seperti Hasrul (2018), Iis et al. (2022), Matdoan & Van Delsen (2020), Paemboan & Abduh (2021), dan lain sebagainya. Berdasarkan beberapa penelitian sebelumnya yang menggunakan analisis *cluster* hierarki *single linkage*, *output* dari penelitian ini nantinya diharapkan dapat menghasilkan data pengelompokan provinsi di Indonesia sesuai *cluster*-nya masing-masing, menggunakan data indikator IPM Indonesia tahun 2023.

II. METODE

A. Metode Pengumpulan Data

Data yang digunakan dalam penelitian adalah data sekunder yang didapatkan dari *website* resmi Badan Pusat Statistik (BPS) terkait Indikator Indeks Pembangunan Manusia (IPM) Indonesia tahun 2023. Data terdiri dari 4 variabel yang akan dilakukan pengelompokan menggunakan metode hierarki *single linkage*. Variabel-variabel tersebut adalah sebagai berikut:

1. Umur Harapan Hidup Saat Lahir (AHH)

Umur Harapan Hidup Saat Lahir adalah perkiraan rata - rata lama hidup seorang bayi yang lahir pada tahun tertentu jika angka kematian pada saat itu tidak berubah sepanjang hidupnya. AHH adalah salah satu ukuran kesehatan masyarakat yang menunjukkan tingkat kesejahteraan dan kualitas hidup suatu kelompok penduduk. AHH diperoleh dengan menggunakan tabel mortalitas yang dibuat berdasarkan data kematian dan jumlah penduduk menurut kelompok usia. Rumus untuk menghitung AHH adalah:

$$AHH = \frac{\sum (\text{Jumlah individu pada setiap kelompok umur} \times \text{Angka harapan hidup pada kelompok umur tersebut})}{\text{Jumlah populasi total}}$$

2. Harapan Lama Sekolah (HLS)

Harapan Lama Sekolah adalah perkiraan rata - rata lama sekolah yang dapat diikuti oleh seorang anak yang berumur 7 tahun dalam sistem pendidikan formal jika pola partisipasi sekolah pada saat itu tidak berubah sepanjang hidupnya. HLS adalah salah satu ukuran pendidikan yang menunjukkan akses dan kesempatan belajar yang tersedia bagi masyarakat. HLS diperoleh dengan menggunakan data angka partisipasi

sekolah menurut jenjang pendidikan dan kelompok usia. Rumus untuk menghitung HLS adalah:

$$HLS = \frac{\sum(\text{Jumlah individu pada setiap kelompok umur} \times \text{Harapan lama sekolah pada kelompok umur tersebut})}{\text{Jumlah populasi total}}$$

3. Pengeluaran per Kapita Disesuaikan (PKD)

Pengeluaran per kapita disesuaikan adalah nilai pengeluaran riil per kapita yang disesuaikan dengan daya beli relatif antar negara atau daerah. PKD adalah salah satu ukuran standar hidup layak yang menunjukkan kemampuan ekonomi dan kesejahteraan material masyarakat. PKD diperoleh dengan menggunakan data produk domestik bruto (PDB) atau produk nasional bruto (PNB) per kapita yang dikonversi ke dolar Amerika Serikat (AS) dengan menggunakan paritas daya beli (PDB). Rumus untuk menghitung PKD adalah:

$$PKD = \frac{\text{Total pengeluaran rumah tangga}}{\text{Jumlah individu dalam rumah tangga}}$$

4. Rata - Rata Lama Sekolah (RLS)

Perkiraan rata - rata lama sekolah yang telah dijalani oleh penduduk berusia 25 tahun ke atas dalam sistem pendidikan formal. RLS adalah salah satu ukuran pendidikan yang menunjukkan tingkat pencapaian dan kualitas pendidikan yang telah diperoleh oleh masyarakat. RLS diperoleh dengan menggunakan data distribusi penduduk berusia 25 tahun ke atas menurut jenjang pendidikan tertinggi yang telah diselesaikan. Rumus untuk menghitung RLS adalah:

$$RLS = \frac{\sum(\text{Jumlah individu pada setiap kelompok umur} \times \text{Rata-rata lama sekolah pada kelompok umur tersebut})}{\text{Jumlah populasi total}}$$

B. Metode Analisis Data

Analisis *cluster single linkage* adalah metode yang akan diimplementasikan pada penelitian ini menggunakan *software* RStudio. Langkah-langkahnya adalah sebagai berikut:

1. Melakukan standarisasi data

Perbedaan skala yang besar dapat mempengaruhi hasil dari perhitungan jarak antar objek menjadi tidak valid sehingga perlu dilakukan standarisasi data. Rumus standarisasi data adalah sebagai berikut (Harnanto, Rusgiyono, & Wuryandari, 2017):

$$Z = \frac{x_i - \bar{x}}{S_x}$$

dimana:

Z = nilai z-score

x_i = nilai pengamatan ke- i

$\bar{x}$ = nilai means

S_x = standart deviasi

2. Melakukan uji asumsi multikolinearitas

Analisis klaster harus memenuhi asumsi bahwa data tidak terjadi korelasi antar variabel sehingga perlu dilakukan uji multikolinieritas untuk mengetahui ada atau tidaknya korelasi. Salah satu cara uji asumsi multikolinieritas adalah menghitung koefisien korelasi *pearson* antar variabel bebas. Apabila terdapat nilai yang lebih dari 0,8 berarti terdapat korelasi antar variabel. Berikut merupakan rumus koefisien korelasi (Hanada & Yanti, 2021):

$$r_{x_1 x_2} = \frac{n \sum_{i=1}^n x_{i1} x_{i2} - (\sum_{i=1}^n x_{i1})(\sum_{i=1}^n x_{i2})}{\sqrt{[n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2][n \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i2})^2]}}$$

Dimana:

r = koefisien korelasi pearson antara dua variabel x_1 dan x_2

n = banyaknya pengamatan

x_{1i} = nilai pengamatan ke- i pada variabel x_1

x_{2i} = nilai pengamatan ke- i pada variabel x_2

$i = 1, 2, 3, \dots, n$

3. Menentukan nilai k optimal dengan metode *silhouette*

Penentuan jumlah *cluster* merupakan salah satu langkah penting dalam analisis *cluster*. Pemilihan jumlah *cluster* didasarkan pada suatu keadaan bahwa jika jumlah kelompok yang dipilih terlalu sedikit akan menimbulkan keragaman dalam kelompok yang besar, namun jika jumlah kelompok yang akan dipilih terlalu besar. Beberapa pedoman yang digunakan dalam menentukan jumlah *cluster* adalah berdasarkan dendrogram dan rasio antar total varian (Putriana, Setyawan, & Noeryanti, 2016).

Nilai *silhouette* dari masing-masing *cluster*, atau bahkan keseluruhan *cluster*, dapat dihitung dari hasil kerja algoritma clustering. Nilai *silhouette* untuk keseluruhan data dengan jumlah *cluster* k dikenal sebagai $sil(k)$, yang didapat dari nilai *silhouette* rata-rata untuk masing-masing *cluster*. Ini memberikan informasi tentang kualitas hasil clustering. Untuk mencari nilai *silhouette* dapat digunakan persamaan sebagai berikut (Paemboan & Abduh, 2021):

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(ci)$$

4. Menghitung jarak antar data dan clustering dengan metode *single linkage*

Jarak yang sering digunakan dalam pengklasteran adalah *Euclidean*. Jarak *Euclidean* harus memenuhi sebuah asumsi yaitu tidak terjadi korelasi antar variabel serta variabel harus memiliki satuan yang sama. Pada metode ini pengukuran jarak dilakukan dengan mengukur akar kuadrat dari penjumlahan kuadrat selisih dari nilai masing-masing obyek. Rumus jarak *Euclidean* adalah sebagai berikut (Windasari, 2020):

$$d(x, y) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Penentuan metode analisis *cluster* merupakan pengelompokan obyek yang dapat dilakukan dengan menggunakan dua macam metode, yaitu metode hierarki dan non-hierarki. Metode hierarki merupakan metode pengklasteran yang terstruktur dan bertahap dimana jumlah kelompoknya belum diketahui. Sedangkan pada metode non-hierarki telah ditentukan jumlah kelompoknya terlebih dahulu. Pada penelitian ini menggunakan metode hierarki yaitu *single linkage* (Windasari, 2020).

Metode *single linkage* merupakan salah satu metode hierarki yang *clusternya* terbentuk berdasarkan jarak terdekat atau similaritas maksimum antara dua data. Penentuan jumlah *cluster* berpedoman pada output *agglomeration* atau dendrogram. Berikut merupakan rumus dari *single linkage* (Iis, et al., 2022):

$$d_{(ij)k} = \min(d_{(i)k}, d_{(j)k})$$

Dimana:

- $d_{(i)k}$: Jarak antar kelompok i dan k
 $d_{(j)k}$: Jarak antar kelompok j dan k
 $d_{(ij)k}$: Jarak antar kelompok ij dan kelompok ke-k

5. Menghitung nilai koefisien korelasi *chopenetic*

Koefisien korelasi *chopenetic* merupakan koefisien korelasi antar elemen asli matriks ketidakmiripan dan elemen-elemen yang dihasilkan dendogram. Apabila nilai koefisien korelasi semakin mendekati 1 maka metode yang digunakan semakin baik. Berikut merupakan rumus koefisien korelasi *cophenetic* (lis, et al., 2022):

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{cik} - \bar{d}_c)}{\sqrt{\left[\sum_{i < k} (d_{ik} - \bar{d})^2 \right] \left[\sum_{i < k} (d_{cik} - \bar{d}_c)^2 \right]}}$$

Dimana:

- r_{coph} : Koefisien korelasi chopnetic
 d_{ik} : Jarak Euclidean objek ke-i dan ke-k
 $\bar{d}$: Rata - Rata d_{ik}
 d_{cik} : Jarak chopenetic objek ke-i dan ke-k
 $\bar{d}_c$: Rata - Rata d_{cik}

6. Melakukan interpretasi dari hasil *clustering* dan menarik kesimpulan

Tahapan interpretasi *cluster* adalah untuk mencari karakter setiap kelompok yang khas, salah satunya dapat dilakukan dengan membandingkan *mean* pada masing-masing kelompok. Selain itu juga dilakukan untuk menjelaskan karakteristik dari setiap kelompok berdasarkan dengan tujuan untuk memberi label pada masing-masing kelompok tersebut (Hasrul, 2018).

III. HASIL DAN PEMBAHASAN

Pada bagian ini akan dibahas pembahasan mengenai analisis pengklasteran pada indikator Indeks Pembangunan Manusia di seluruh Indonesia.

1. Data

Data yang digunakan pada penelitian ini adalah empat indikator Indeks Pembangunan Manusia di seluruh provinsi Indonesia pada tahun 2023. Berikut merupakan data yang diperoleh:

Tabel 1. Data IPM Indonesia Tahun 2023

Provinsi	X ₁	X ₂	X ₃	X ₄
Aceh	70.3 4	14.3 8	1033 4	9.55
Sumatera Utara	69.9 8	13.4 8	11049	9.82
Sumatera Barat	70.1 9	14.1 1	11380	9.28
Riau	72.2 4	13.3 0	11448	9.32

Jambi	71.7 7	13.1 3	11160	8.81
Sumatera Selatan	70.6 6	12.6 3	11472	8.50
Bengkulu	69.9 2	13.7 4	11172	9.03
Lampung	71.2 5	12.7 7	1076 9	8.29
Kep. Bangka Belitung	71.2 3	12.3 1	1358 9	8.25
Kepulauan Riau	70.9 1	13.0 5	1499 8	10.4 1
DKI Jakarta	73.6 5	13.3 3	1937 3	11.4 5
Jawa Barat	73.8 0	12.6 8	11695	8.83
Jawa Tengah	74.6 9	12.8 5	11835	8.01
DI Yogyakarta	75.1 2	15.6 6	1492 4	9.83
Jawa Timur	72.1 1	13.3 8	1242 1	8.11
Banten	70.7 7	13.0 9	1260 1	9.15
Bali	72.9 8	13.5 8	1438 2	9.45
Nusa Tenggara Barat	67.4 7	13.9 7	11095	7.74
Nusa Tenggara Timur	67.7 7	13.2 2	8248	7.82
Kalimantan Barat	71.3 2	12.6 7	9810	7.71
Kalimantan Tengah	70.2 7	12.7 6	11878	8.73
Kalimantan Selatan	69.4 2	12.8 6	1295 3	8.55
Kalimantan Timur	74.7 2	14.0 2	1320 2	9.99
Kalimantan Utara	72.6 9	13.2 0	9734	9.34
Sulawesi Utara	72.4 0	12.9 6	11497	977

Sulawesi Tengah	69.1 7	13.3 3	1014 9	8.96
Sulawesi Selatan	71.2 2	13.5 4	11841	8.76
Sulawesi Tenggara	71.4 7	13.7 0	10117	9.31
Gorontalo	68.8 3	13.1 6	11069	8.10
Sulawesi Barat	66.0 1	12.8 8	9718	8.13
Maluku	66.7 8	14.0 8	9278	10.2 0
Maluku Utara	69.1 1	13.7 4	8834	9.26
Papua Barat	66.7 9	13.3 4	8404	7.93
Papua	66.4 4	11.1 5	7562	7.15

Dimana keterangan variabel pada Tabel 1 adalah:

X_1 = Umur Harapan Hidup Saat Lahir (AHH)

X_2 = Harapan Lama Sekolah (HLS)

X_3 = Pengeluaran per Kapita Disesuaikan (PKD)

X_4 = Rata - Rata Lama Sekolah (RLS)

2. Standarisasi Data

Proses standarisasi data dilakukan ketika data memiliki perbedaan skala besar antara variabel penelitian. Hal tersebut dilakukan karena perbedaan skala yang besar antar variabel dapat mempengaruhi hasil dari perhitungan jarak kemiripan antar objek menjadi tidak valid. Hasil dari standarisasi data adalah sebagai berikut.

Tabel 2. Hasil Standarisasi Data

Provinsi	X_1	X_2	X_3	X_4
Aceh	-0.145893 1	1.4700395	-0.501607 9	0.6824709
Sumatera Utara	-0.295577 2	0.2499904	-0.185985 1	0.9785524
Sumatera Barat	-0.208261 5	1.1040248	-0.039871 6	0.3863895
Riau	0.6441065	0.0059806	-0.009854 3	0.4302534
Jambi	0.4486855	-0.224473 1	-0.136986 3	-0.1290115

Sumatera Selatan	-0.012840 6	-0.902278 1	0.0007400	-0.468956 9
Bengkulu	-0.320524 6	0.6024491	-0.131689 1	0.1122400
Lampung	0.2324751	-0.712492 7	-0.309585 6	-0.699242 4
Kep. Bangka Belitung	0.2241593	-1.336073 3	0.9352485	-0.743106 3
Kepulauan Riau	0.0911068	-0.332921 9	1.5572242	1.6255451
DKI Jakarta	1.2303693	0.0466489	3.4884829	2.7660070
Jawa Barat	1.2927377	-0.834497 6	0.0991791	-0.107079 6
Jawa Tengah	1.6627901	-0.604043 9	0.1609793	-1.006289 8
DI Yogyakarta	1.8415795	3.2052203	1.5245583	0.9895184
Jawa Timur	0.5900539	0.1144294	0.4196576	-0.896630 1
Banten	0.0328963	-0.278697 5	0.4991151	0.2438318
Bali	0.9517905	0.3855514	1.2853030	0.5728111
Nusa Tenggara Barat	-1.339208 3	0.9142394	-0.165679 3	-1.302371 3
Nusa Tenggara Timur	-1.214471 5	-0.102468 2	-1.422432 1	-1.214643 5
Kalimantan Barat	0.2615804	-0.848053 7	-0.732917 5	-1.335269 2
Kalimantan Tengah	-0.174998 4	-0.726048 8	0.1799609	-0.216739 4
Kalimantan Selatan	-0.528419 2	-0.590487 8	0.6544987	-0.414127 0

Kalimantan Timur	1.6752638	0.9820199	0.7644149	1.1649740
Kalimantan Utara	0.8312116	-0.1295804	-0.7664663	0.4521854
Sulawesi Utara	0.7106328	-0.4549268	0.0117758	0.9237225
Sulawesi Tengah	-0.6323665	0.0466489	-0.5832726	0.0354782
Sulawesi Selatan	0.2200014	0.3313270	0.1636279	-0.1838414
Sulawesi Tenggara	0.3239487	0.5482247	-0.5973984	0.4192874
Gorontalo	-0.7737349	-0.1838048	-0.1771565	-0.9075960
Sulawesi Barat	-1.9462606	-0.5633756	-0.7735292	-0.8746981
Maluku	-1.6261029	1.0633565	-0.9677586	1.3952596
Maluku Utara	-0.6573139	0.6024491	-1.1637538	0.3644575
Papua Barat	-1.6219450	0.0602050	-1.3535689	-1.0940177
Papua	-1.7674712	-2.9085810	-1.7252534	-1.9493641

Tabel 2 adalah hasil setelah penyesuaian atau normalisasi skala dari keempat variabel menjadi seragam atau setara, sehingga memudahkan analisis dan interpretasi data.

3. Uji Asumsi Multikolinearitas

Pada tahap ini dilakukan uji asumsi multikolinearitas untuk melihat bagaimana hubungan antar variabel independen Angka Harapan Hidup, Harapan Lama Sekolah, dan Pengeluaran Per Kapita, Rata – Rata Lama Sekolah dengan menghitung nilai korelasi antara dua variabel. Hasil dari perhitungan menggunakan *software* Rstudio dapat disajikan dalam tabel berikut.

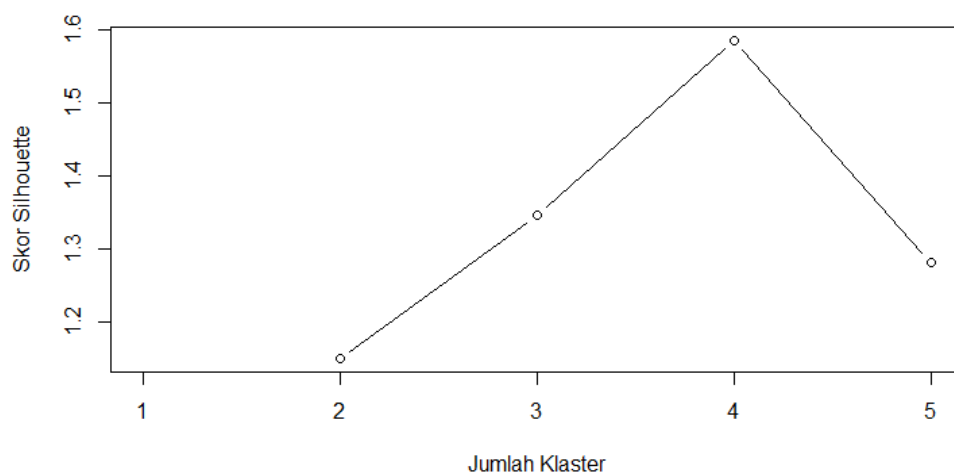
Tabel 3. Hasil Uji Asumsi Multikolinearitas

	Angka Harapan Hidup	Harapan Lama Sekolah	Pengeluaran Per Kapita	Rata – Rata Lama Sekolah
Angka Harapan Hidup	1.0000000	0.2339179	0.6316007	0.4326277
Harapan Lama Sekolah	0.2339179	1.0000000	0.2171719	0.4989544
Pengeluaran Per Kapita	0.6316007	0.2171719	1.0000000	0.5985096
Rata – Rata Lama Sekolah	0.4326277	0.4989544	0.5985096	1.0000000

Berdasarkan Tabel 3. Diketahui variabel Angka Harapan Hidup dengan Pengeluaran Per Kapita yaitu 0.63, dan variabel Pengeluaran Per Kapita dengan Rata – Rata Lama Sekolah yaitu 0,59. Kemudian untuk variabel Angka Harapan Hidup dengan Harapan Lama Sekolah yang bernilai 0,23, Harapan Lama Sekolah dengan Pengeluaran Per Kapita bernilai 0,21, variabel Harapan Lama Sekolah dengan Rata – Rata Lama Sekolah bernilai 0,49, dan variabel Angka Harapan Hidup dengan Rata – Rata Lama Sekolah yaitu 0,43 menunjukkan nilai korelasi yang rendah. Berdasarkan uji asumsi korelasi *pearson*, nilai korelasi yang kurang dari 0,8 menunjukkan tidak terjadi multikolinearitas yang berarti hubungan antara dua variabel tersebut lemah. Sehingga variabel-variabel tersebut dapat diolah untuk dilakukan *clustering*.

0. Metode *Silhouette*

Pada tahap ini dilakukan pencarian nilai k optimal dalam analisis *clustering* untuk menunjukkan jumlah *cluster* yang dianggap optimal atau paling sesuai berdasarkan metrik evaluasi yang dipilih. Metode *silhouette* dilakukan untuk menentukan jumlah *cluster* yang akan digunakan untuk *clustering single linkage* pada penelitian ini. Kemudian dipilih nilai k yang akan diuji yaitu k=1,2,3,4,5. Dengan bantuan *software* Rstudio didapatkan hasil nilai *silhouette* yang direpresentasikan dengan plot pada Gambar 1.

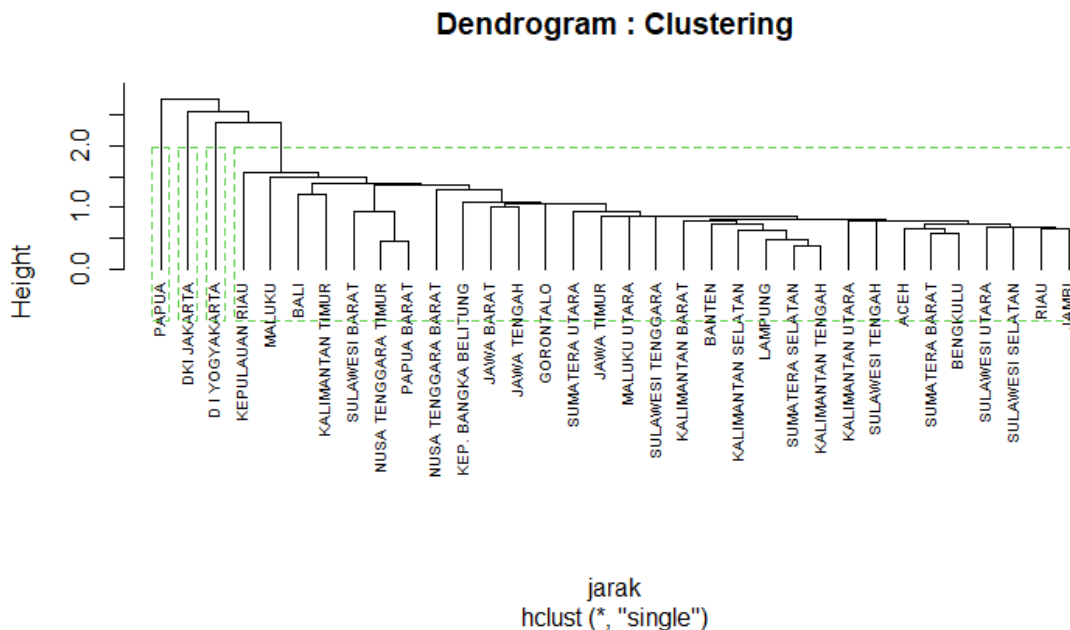


Gambar 1. Plot Nilai *Silhouette*

Berdasarkan Gambar 1. Diketahui nilai *silhouette* optimum adalah pada saat $k=4$. Hasil tersebut menunjukkan bahwa pembagian data ke dalam 4 *cluster* memiliki tingkat kohesi yang lebih tinggi dan pemisahan yang lebih baik antara *cluster*. Dapat diartikan bahwa *clustering single linkage* yang akan dilakukan yaitu menggunakan 4 *cluster*.

0. Clustering Hierarki Single Linkage

Setelah diketahui bahwa k optimal adalah 4, maka selanjutnya dilakukan pencarian jarak antar data menggunakan jarak *euclidean* yang digunakan untuk analisis clustering dengan metode *single linkage* dengan membagi menjadi kelompok dengan 4 *cluster*. Pada tahap ini dilakukan menggunakan *software* Rstudio. Hasil pembagiannya *cluster* dapat direpresentasikan dengan dendrogram pada Gambar 2.



Gambar 2. Plot Dendrogram Hasil *Clustering*

Berdasarkan Gambar 2, ada dua cluster utama yang ditunjukkan oleh garis putus-putus hijau, dengan beberapa sub-cluster di dalamnya. Cluster pertama terdiri dari entitas-entitas seperti Banyuwangi, Jember, Situbondo, dan Bondowoso, yang mungkin memiliki karakteristik yang mirip. Cluster kedua terdiri dari entitas-entitas seperti Malang, Surabaya, Sidoarjo, dan Gresik, yang mungkin memiliki karakteristik yang berbeda dari cluster pertama.

Tinggi garis yang menghubungkan entitas-entitas menunjukkan jarak antara mereka. Semakin tinggi garis, semakin jauh jaraknya. Misalnya, Banyuwangi dan Jember terhubung oleh garis yang sangat pendek, yang berarti mereka sangat mirip. Sedangkan Malang dan Surabaya terhubung oleh garis yang sangat panjang, yang berarti mereka sangat berbeda.

0. Koefisien Korelasi *Chopenetic*

Pada tahap ini dilakukan perhitungan koefisien korelasi *chopenetic* dengan menggunakan *software* RStudio dan didapatkan nilai 0,8335765. Nilai yang didapatkan tersebut adalah nilai yang cukup tinggi yaitu mendekati 1, ini mengindikasikan bahwa dendrogram yang telah dihasilkan secara keseluruhan memberikan gambaran yang baik tentang hubungan jarak antara titik-titik data dalam analisis *clustering* hierarki yang dilakukan. Dendrogram yang dihasilkan dari analisis *clustering* hierarki cukup baik dalam merepresentasikan jarak sebenarnya antara titik data. Maka dapat disimpulkan bahwa

metode hierarki single linkage yang telah dilakukan dapat digunakan pada pengklasteran data dalam penelitian ini dengan baik.

0. Interpretasi Hasil

Berdasarkan hasil clustering menggunakan *software* R, diperoleh anggota setiap cluster sebagai berikut:

Tabel 4. Hasil Pengelompokan Provinsi berdasarkan IPM Indonesia Tahun 2023

Cluster	Anggota
1	Maluku, Bali, Sumatera Barat, NTT, Papua Barat, NTB, Kep. Bangka Belitung, Jawa Barat, Jawa Tengah, Gorontalo, Sumatera Utara, Jawa Timur, Maluku Utara, Sulawesi Tenggara, Kalimantan Barat, Banten, Kalimantan Selatan, Lampung, Sumatera Selatan, Kalimantan Tengah, Kalimantan Utara, Sulawesi Tengah, Aceh, Sumatera Barat, Bengkulu, Sulawesi Utara, Sulawesi Selatan, Riau, Jambi, Kepulauan Riau
2	DI Yogyakarta
3	DKI Jakarta
4	Papua

Dengan karakteristik setiap *cluster*:

Tabel 5. Karakteristik Setiap *Cluster*

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
X ₁ (Angka Harapan Hidup)	70,59	13,29	11230,06	8,87
X ₂ (Harapan Lama Sekolah)	73,65	12,33	19373	11,45
X ₃ (Pengeluaran Per kapita)	75,12	15,66	14924	9,83
X ₄ (Rata - Rata Lama Sekolah)	66,44	11,15	7562	7,15



Gambar 3. Hasil Pengelompokan Provinsi berdasarkan IPM Indonesia Tahun 2023

Keterangan pada Gambar 3 adalah sebagai berikut:

- = Klaster 1
- = Klaster 2
- = Klaster 3
- = Klaster 4

Berdasarkan Tabel 4. Dan Tabel 5. diperoleh 4 kategori klaster, yaitu:

1. Klaster 1 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang cukup baik dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 31 Provinsi, yaitu Maluku, Bali, Sumatera Barat, NTT, Papua Barat, NTB, Kep. Bangka Belitung, Jawa Barat, Jawa Tengah, Gorontalo, Sumatera Utara, Jawa Timur, Maluku Utara, Sulawesi Tenggara, Kalimantan Barat, Banten, Kalimantan Selatan, Lampung, Sumatera Selatan, Kalimantan Tengah, Kalimantan Utara, Sulawesi Tengah, Aceh, Sumatera Barat, Bengkulu, Sulawesi Utara, Sulawesi Selatan, Riau, Jambi, Kepulauan Riau.
2. Klaster 2 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang baik dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 1 Provinsi, yaitu DI Yogyakarta.
3. Klaster 3 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang kurang dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 1 Provinsi, yaitu DKI Jakarta.
4. Klaster 4 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang sangat kurang dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 1 Provinsi, yaitu Papua.

IV. SIMPULAN

Berdasarkan analisis *clustering* dengan metode hierarki *single linkage* terhadap Indeks Pembangunan Manusia (IPM) Indonesia tahun 2023 yang telah dilakukan pada penelitian ini, didapatkan kesimpulan bahwa pengelompokan provinsi di Indonesia terbagi menjadi empat *cluster*, yaitu:

1. Klaster 1 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang cukup baik dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 31 Provinsi.
2. Klaster 2 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang baik dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 1 Provinsi.
3. Klaster 3 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang kurang dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 1 Provinsi.
4. Klaster 4 menunjukkan bahwa Provinsi di Indonesia memiliki karakteristik yang sangat kurang dalam variabel X_1 (Angka Harapan Hidup), X_2 (Harapan Lama Sekolah), X_3 (Pengeluaran Per Kapita), X_4 (Rata - Rata Lama Sekolah). Terdiri dari 1 Provinsi.

Selain itu, didapatkan nilai koefisien korelasi *chopenetic* yaitu 0,8335765 yang menunjukkan bahwa metode hierarki *single linkage* adalah metode yang baik untuk analisis *cluster* pada penelitian ini. Diharapkan penelitian yang telah dilakukan dapat bermanfaat bagi pemerintah atau pihak terkait untuk menjadi bahan evaluasi terkait IPM di Indonesia di masa yang akan datang, serta bermanfaat dalam membantu pemerintah dalam merancang kebijakan pembangunan negara yang lebih tepat sasaran dan lebih efektif sesuai kebutuhan setiap wilayah.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2023). Diambil kembali dari BPS: <https://www.bps.go.id/id>
- Hanada, S., & Yanti, T. S. (2021). Penggunaan Analisis Cluster dalam Pengelompokan Kecamatan di Kabupaten Karawang Berdasarkan Metode Kontrasepsi Peserta KB Aktif. *Prosiding Statistika*, 7(1), 42-49.
- Harnanto, Y. I., Rusgiyono, A., & Wuryandari, T. (2017). Penerapan Analisis Klaster Metode Ward Terhadap Kabupaten/Kota di Jawa Tengah Berdasarkan Pengguna Alat Kontrasepsi. *Jurnal Gaussian*, 6(4), 528-537.
- Hasrul, M. (2018). *Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat*. Berdasarkan Indikator Kesejahteraan Rakyat Makassar, Makassar.
- Iis, Yahya, I., Wibawa, G. N., Baharuddin, Ruslan, & Laome, L. (2022). Penggunaan Korelasi Cophenetic untuk Pemilihan Metode Cluster Berhierarki pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020. *Prosiding Seminar Nasional Sains Dan Terapan (Sinta) Vi*, 1-16.
- Matdoan, Muhammad Yahya, and Marlon Stivo Noya Van Delsen. "Penerapan Analisis Cluster dengan Metode Hierarki untuk Klasifikasi Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia." *STATMAT: Jurnal Statistika dan Matematika* 2.2 (2020): 1-10.
- Paemboan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat. *PENA TEKNIK : Jurnal Ilmiah Ilmu - Ilmu Teknik* 6(2), 48-54.
- Putriana, U., Setyawan, Y., & Noeryanti. (2016). Metode Claster untuk Analisis Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Variabel yang Mempengaruhi Kemiskinan pada Tahun 2013. *Jurnal Statistika Industri dan Komputasi*, 1(1), 38-52.
- Rencher, A. C. (2002). *Methods of multivariate analysis* (2nd ed.). Wiley.
- Widodo, E., Mashita, S., & Prasetyowati, Y. (2020). Perbandingan Metode Average Linkage, Complete Linkage, dan Ward's pada Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia. *Faktor Exacta*, 13(92), 81-87.
- Windasari, R. (2020). *Analisis Cluster Hirarki Metode Average Linkage Berdasarkan Jumlah Kriminalitas di Indonesia Tahun 2019*. Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang.
- Wulansari, A. (2012). Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Variabel-Variabel Indeks Pembangunan Manusia. *eL-Qist*, 2(2), 367-386.